Visualization and Interaction for Knowledge Discovery in Simulation Data

Niclas Feldkamp
Sören Bergmann
Steffen Strassburger
Technische Universität Ilmenau
{niclas.feldkamp, soeren.bergmann,
steffen.strassburger}@tu-ilmenau.de

Thomas Schulze
Otto-von-Guericke-Universität Magdeburg
tom@isg.cs.uni-magdeburg.de

Abstract

Discrete-event simulation is an established and popular technology for investigating the dynamic behavior of complex manufacturing and logistics systems. Besides traditional simulation studies that focus on single model aspects, data farming describes an approach for using the simulation model as a data generator for broad scale experimentation with a broader coverage of the system behavior. On top of that we developed a process called knowledge discovery in simulation data that enhances the data farming concept by using data mining methods for the data analysis. In order to uncover patterns and causal relationships in the model, a visually guided analysis then enables an exploratory data analysis. While our previous work mainly focused on the application of suitable data mining methods, we address suitable visualization and interaction methods in this paper. We present those in a conceptual framework followed by an exemplary demonstration in an academic case study.

1. Introduction

In the context of manufacturing and logistics, simulation is a well-established tool for planning, operating, and monitoring complex systems. Traditionally, simulation studies have been used to account for project goals formulated beforehand such as optimizing a buffer capacity or deciding between multiple shop floor layouts [22]. In order to achieve those goals, the simulation study is carried out by comparing predetermined scenarios that the user already had in mind before. There are cases where automated simulation-based optimization algorithms are used, but even then, the target function has to be set up beforehand. Therefore, the simulation analyst usually takes an educated guess based on his experience, which input parameters (factors) might be influential on the project scope [21]. Kleijnen describes this as the trial-and-error approach for finding good solutions and that simulation users should spent more time analyzing their models or model output, respectively. Besides, the development of simulation models is a very time-consuming and costly process, so that the potential of the model should be leveraged as much as possible [21].

With the widespread availability of Big Data infrastructures and ease of access to computing power, there are new possibilities for leveraging the potential of simulation models. In this regard, we developed a concept called Knowledge Discovery in Simulation Data that allows to use a combination of data mining and visual analysis in order to find hidden and potentially interesting and useful knowledge in the system outside of prior defined project scopes [8]. The process can support the decision-making in a manufacturing context that may lead to outside-of-the-box solutions that the analyst possibly did not think of before. This process is based on covering the whole range of possible system behavior through large-scale experiments. These large amounts of simulation data can then be analyzed using data mining algorithms. Besides the proof of the general applicability through various case studies [10, 11, 35], our previous research mainly focused on the computational side regarding suitable data mining methods.

In this paper, we want to focus on suitable interaction and visualization methods. The remainder of this paper is structured as follows: In section 2, we give a short overview over the related work. Section 3 introduces the general process of knowledge discovery in simulation data followed by the actual discussion of interaction and visualization possibilities. In section 4, we provide a short academic case study to demonstrate the process. Section 5 gives some concluding remarks and an outlook to future work.

URI: https://hdl.handle.net/10125/63904
978-0-9981331-3-3

2. Related Work

2.1. Data Farming

Originally developed for military and warfare simulations, data farming is an approach for using the simulation model as a data generator [6, 15, 29]. In order to cover the whole bandwidth of possible system behavior, the simulation model has to be run many times with different variants of input factor parameter sets. Because the number of possible input sets grows exponentially with the number of factors k and the number of factor values n, a naïve, brute force n^k approach is not feasible for large factor spaces. Therefore, one major effort of the data farming research discipline is the implementation of more efficient experimental designs, like for example the nearly orthogonal Latin hypercube (NOLH) [5, 37]. Those can reduce the number of experiments dramatically while still maintaining a good coverage of the models response surface as well as other desirable features of a good experimental design like balanced factor values and orthogonality for bias-free analysis [30, 31]. The farming metaphor describes the maximization of data output in the most efficient way, resembling a farmer who cultivates his land in order to maximize his crop yield [31].

Nonetheless, in complex models with lots of factors, very large numbers of experiments are yet to be expected. Therefore, high performance computing and parallelization of experiments is a crucial requirement. In order to analyze the generated massive amount of simulation data, some form of automated analysis is needed. To enhance this process, we developed a concept called knowledge discovery in simulation data for using data mining and visualization methods that can be applied on massive scale farmed simulation data, integrating the aspects of data farming, data mining and visual analytics [8].

2.2. Visual Analytics

Visual analysis in general is an important tool when a human interpretation of data is required. In traditional simulation studies, the most commonly used techniques include animation of process flow, time plots, and graphs of selected outputs [40]. Visual analytics on the other hand is a discipline of its own going beyond commonly used visualization methods in simulation analysis. Visual analytics describes a human-in-the-loop process between automated data analysis assisted by data mining algorithms and the corresponding visualization of results in an interactive manner [36]. Therefore, visual analytics can be defined as an "iterative process that involves information gathering, data preprocessing, knowledge representation, interaction and decision making" [19]. The goal here is that the user can switch between modifying the data mining algorithms and interacting with the visualization in order to build up knowledge and draw conclusion from it.

This approach is advantageous because the human mind is able to identify patterns and relations in visual representations quickly.

Due to the tight integration of visualization and data mining, the visual analytics approach is generally very useful for knowledge discovery. The commonly known process for knowledge extraction is called knowledge discovery in databases (KDD), which describes how to derive and transform data into knowledge from heterogeneous sources [7]. The logical advancement of this process is to enhance it with a component of interactive, visual exploration in order to benefit from human interpretation and expert background knowledge [26]. This has been done in many applications and for different kinds of data, e.g. in bioinformatics and biomedical data [20, 26], text mining [28], or movement patterns in GPS data [17]. Especially unsupervised data mining operations benefit from appropriate visual exploration and interaction since pattern detection and investigation of multidimensional data is one of the main goals of knowledge discovery [3]. This has been done for example for the visual exploration of web clickstream data [39]. The concept of VA and VA-related applications have also gained importance in the era of big data in order to make use of large data collections and extreme-scale data [1], e.g. weather and climate data [34]. On the other hand, VA has had influence on computational methods in terms of improving their performance and scalability for real-time visualization of big data [4].

Still, data mining in general and VA in particular are not very prevalent in the simulation community and research regarding simulation output analysis. Recently, there have been some papers on using VA for the analysis of stochastic simulations of particle movement and chemical phenomena [23] and complex physical systems [24, 33], where output data can exhibit complex characteristics and is therefore difficult to analyze and to interpret.

However, the usage of visual analytics in conjunction with the analysis and knowledge discovery aspect of data farming and especially for the analysis of discrete-event manufacturing simulations is a rather new topic. In fact it is perfectly suitable for analyzing large amounts of simulation data that have been generated by data farming in order to find hidden relations and conclusions about the modelled system.

Therefore, simulation analysis can benefit from the body of experience that has been achieved in VA-research. In our previous work, we focused primarily on the elaboration of suitable data mining algorithms and data analysis methods, as well as the validation of the practical applicability through real world case studies. In this paper, we focus on suitable visualization and interaction possibilities.

3. Knowledge Discovery in Simulation Data

3.1. General Process

The general process for knowledge discovery in simulation data is shown in Figure 1. This process is separated into two main areas. One is for data generation (green area) and one is for data analysis (blue area). The process starts with the definition, distribution, and execution of simulation experiments. On an abstract level, the simulation model itself acts as a black box that simply transforms a set of factors into a corresponding set of outputs.

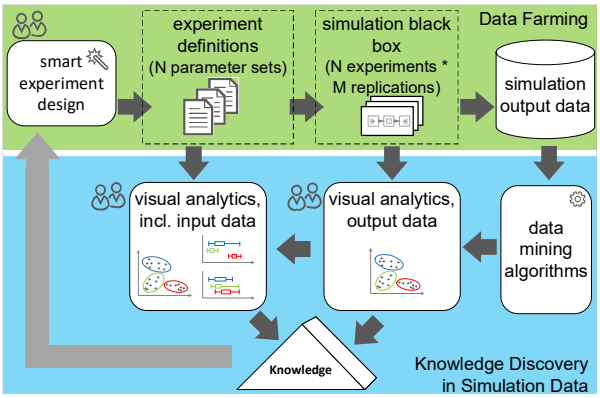

Figure 1. Process for knowledge discovery in simulation data [8].

The analysis then starts on the output data. By applying unsupervised data mining methods like clustering algorithms or Gaussian mixture modelling, multidimensional patterns in the outputs can be detected. This allows grouping the behavior of the simulation model into multidimensional categories. With the use of suitable visualizations, these groups can then be categorized and interpreted.

In the next step, factors can be added to the process by linking them visually to the priorly created groups of output dimensions. Additionally, the analysis of the relations between input and output data can be assisted by supervised data mining methods like linear regression, logistic regression, classification trees, correlation tables and association rule mining which all need specialized forms visual representations. Therefore, the next section discusses possible visualization and interaction possibilities.

3.2. Visualization and Interaction Methods

The visualization of farmed simulation data is a challenging task, because the data is large, multidimensional, and multivariate. Multivariate means, that multiple factors and multiple result parameters need to be considered in the data analysis to determine the system behavior. The most important methods for the visualization of multidimensional data are scatterplot matrices, parallel coordinate plots, and spider charts [16, 27, 38]. For the visualization of single parameters, boxplots and histograms are suitable [13]. Commonly used visualizations for categorical parameters are pie charts and bar charts [25]. For some data mining methods, specialized visualization methods need to be considered, like flowcharts for classification trees and graphs for Bayesian networks [12]. Figure 2 gives an overview of data mining methods and suitable visualization methods for data farming analysis.

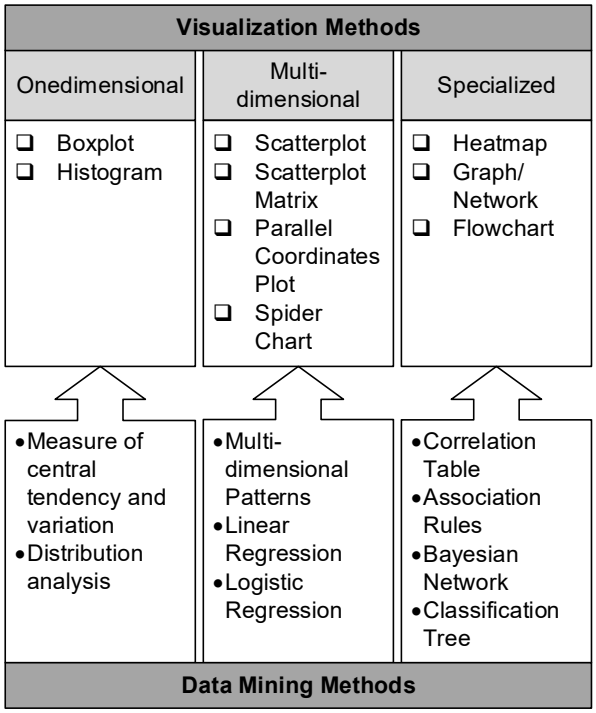

Figure 2. Matching of data mining methods with suitable visualization methods.

Regarding interaction methods, there are distinct operations of how users can interact with the data including selection, navigation (zooming and browsing), arranging, and filtering [18, 40]. Those basic interaction operations can then be combined in order to fulfill a certain visual task. Those visual tasks have various levels of complexity. From identifying a certain object to finding a relation between objects to finding patterns and structures to group and classify objects based on their perceived features, these tasks require an increased cognitive effort [2]. On the other hand, abstract analysis objectives can be used to frame the analysis process for knowledge discovery in simulation data. These objectives are derived from the original data farming guidelines for warfare applications and have been generalized for a wider applicability [14]:

- How are output values distributed?
- Which outputs are relevant? Do structures, patterns, and correlations exist within the data?
- How are input factors values distributed among selected patterns of system behavior?
- Which factors are most influential on the system behavior?
- Are there any significant interactions between factors?
- Which outputs are sensitive to stochastic behavior in the system and to what extent?

Beyond that, analysis objectives can be extended with more specific questions that are tailored to the underlying model and its context. The aforementioned analysis objectives can then be processed by building visual tasks from combining data mining, visualization, and interaction methods. Table 1 gives an overview on the analysis objectives with corresponding visual tasks. Note that the visual tasks given in Table 1 reflect the aforementioned general analysis process for knowledge discovery in simulation data shown in Figure 1 that starts on the output data and continues to include selected input factors.

We start by analyzing the distribution and variance of outputs, e.g. by identifying peaks, dips, and generally interesting trends in their respective value distributions. By using unsupervised data mining, we can then try to find multidimensional patterns and classify those in order to create disjoint groups of output behavior. In the next step, we add input factors to the visualization in order to analyze their relation to the priorly analyzed outputs. Supervised data mining like regression and classification trees can be incorporated to further support the visual analysis.

Table 1. Analysis objectives and corresponding visual tasks

Analysis Objective	Visual Task
How are output values distributed?	- Identify distributions, peaks, and dips. - Comparing to other parameters
Which outputs are relevant? Do structures, patterns, and correlations exists within the data?	Localization of equally distributed outputs - Finding relationships through correlation - Association of indirect relations through interpretation of context - Finding multi-dimensional structures and patterns - Associating indirect relations through comparison of visual patterns - Classifying through comparisons of structures among groups
How are factors values distributed among selected patterns of system behavior?	- Identify distributions in selected subsets - Associating relations through comparisons of visual patterns
Which factors are most influential on the system behavior?	- Localization of clusters and finding of relationships by adding factors to the visualization - Comparing of clusters - Comparing regression lines
Are there any significant interactions between factors?	- Comparing regression lines among different input/output value combinations
Which outputs are sensitive to stochastic behavior in the system and to what extent?	- Identify and compare distribution of outputs that are exposed to stochastic influence among fixed factor settings

In the next section, we demonstrate the process of knowledge discovery in simulation data using a simple academic case study. Some exemplary visualization

and interaction methods are selected in order to generate knowledge about the system behavior.

4. Case Study

In this section, we demonstrate the process of knowledge discovery in simulation by means of a simple single server model. This model has been implement in Siemens Plant Simulation. Figure 3 shows a screenshot of the simulation model.

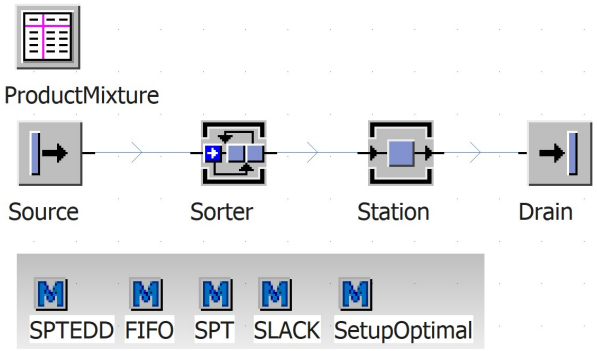

Figure 3. Screenshot of the simulation model.

In this model, seven different types of products can enter the system from the source into a sorter that acts as a queue that can resort the products based on a selected sorting strategy. After sorting, they enter the actual station in order to be processed. The implemented sorting strategies include first in first out (FIFO), shortest processing time (SPT), minimum slack time (SLACK), a weighted combination of SPT and earliest due date (SPTEDD), and sorting according to current station setup state (SetupOptimal). The product types differ in process and setup times on the station and in their designated due date. Table 2 shows input factors and their value limits for which we want to define simulation experiments.

Table 2. Input factors.

Input factor	Margins
Interarrival time of products at the source	60s-240s
Sorter capacity	10-1000 slots
Sorter strategy	5 strategies
Product mixture (7 product types)	0-100% per product

For interarrival times of products, fixed intervals from 60 seconds up to 240 seconds were defined with increments of 10 seconds, so that there are 19 different levels for interarrival time. The capacity of the sorter enumerates from 100 to 1000 slots in increments of 100, because preliminary experiments showed that a sorter allocation over 1000 is not expectable, even with a high arrival frequency. In addition, the five implemented sorting strategies represent a factor. Given this three factors, there are 19*10*5 = 950 different combinations of factor values in a full factorial experimental design. We also aimed to incorporate the mixture of the seven product types as a factor, which means that there is one factor per product describing each products proportion in the mix. Each proportion needs to be varied between 0% and 100%, which results in way too many possible combinations when using a full factorial design. We therefore used a sampling technique that is very common in data farming called Nearly Orthogonal Latin Hypercube Sampling (NOLHS), which can dramatically reduce the number of experiments while maintaining a good coverage of the input space. By using NOLH-Sampling, we created 512 different product mixtures. Multiplied with the other factor value combinations this results in a final experimental design with 486,400 experiments.

After the experiments were conducted, the data analysis starts with the output data. The input factor values are balanced and equally distributed, which is a desired feature of good experimental design in order to avoid bias. Because of that, patterns in the output data need to be detected first. If we create subsets from these patterns or any other manually applied filter, corresponding factor values that belong to the remaining experiments in that subset exhibit a skewed distribution. The more skewed a factor value distribution is, the more dominant and therefore decisive this factor value is for the given filter setting or experiment subset respectively. This method is called skewed distribution analysis [32]. Because our filters on the data come from multidimensional patterns, we can therefore determine the relation of systems behavior in terms of multidimensional patterns to its factors.

Figure 4 shows a scatterplot matrix for five selected output parameters. This matrix shows 2D-scatterplots for all combinations of the selected output dimensions. Diagonally from top left to bottom right, it shows histograms for each distribution of parameter values. As we can see, sorter utilization and cycle time are equally distributed among the other parameters. We can also see that their respective histograms show a very one-sided peak on the left. Because these two parameters do not exhibit much variation, they are excluded from further analysis because they will not reveal interesting patterns and bias the meaningfulness of data mining results.

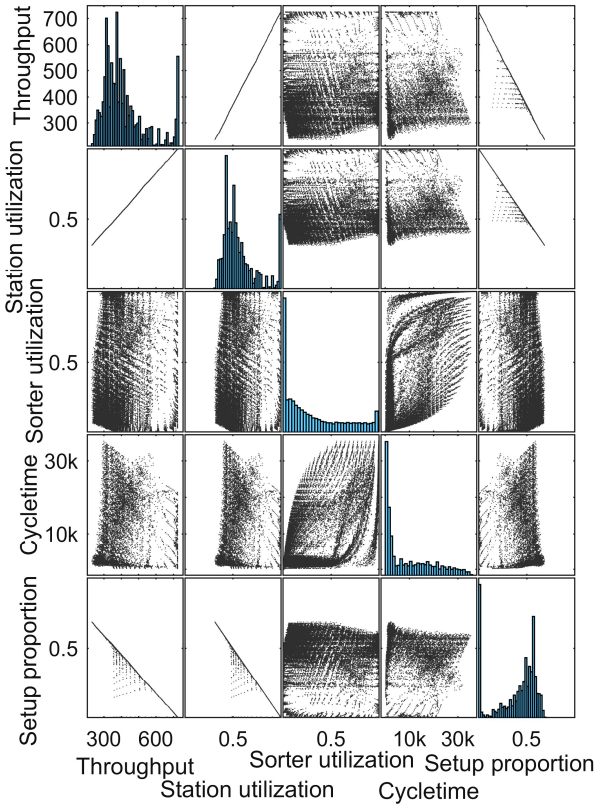

Figure 4. Scatterplot matrix for relevant output dimensions that shows every combination of 2D-scatterplots. Additionally, histograms for the value distribution of each output have can be seen on the diagonal axis.

We then applied a clustering algorithm (k-means clustering with 5 clusters) on the remaining parameters throughput, station utilization, and percentage of setup processes (setup proportion). The clustering algorithm groups the simulation experiments based on their similarity. As a result, experiments within the same cluster are similar regarding the selected output performance measures, and experiments in different clusters are very dissimilar.

A visualization of the clustering results is shown in Figure 5 by means of a parallel coordinate plot. In this plot, each vertical axis represents one of the three dimensions that has been used for the clustering and each horizontal line represents one single simulation experiment. The cluster allocation of each experiment (Cluster 0-4) was added as an additional axis. Although this parameter is not metrical, this axis can be used to filter the data among the clusters for a more usability convenience.

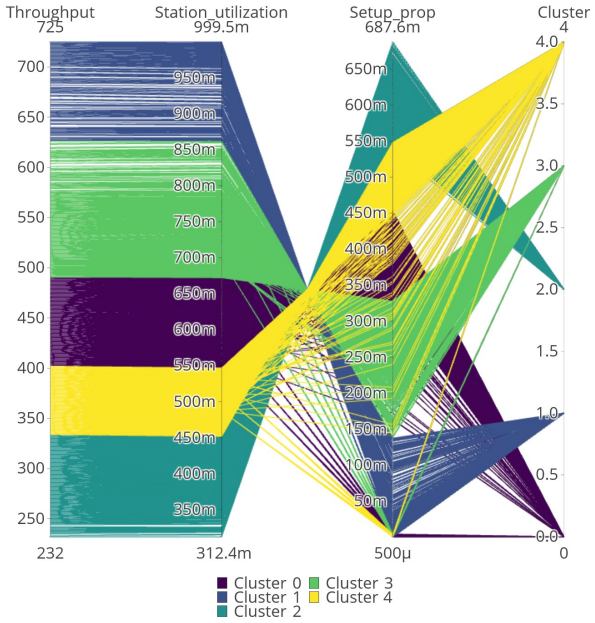

Figure 5. Parallel coordinate plot for selected outputs and corresponding clusters by color. Each vertical axis represents an output dimension and each horizontal line represents one simulation experiment.

Figure 5 shows that there are patterns in the output dimensions that allow to structure and group the system behavior in a qualitative manner. Throughput and station utilization are strongly correlated with each other which allows to group the system behavior in good performance and bad performance accordingly, given that a high throughput and station utilization are considered to be desirable. Setup proportion is less clearly separable with lots of outliers in the lower end. Assuming that a low percentage of setup processes is desirable because they are not value adding, we can define cluster 1 (blue) to be composed of experiments with the good system performance and cluster 2 (cyan) as the cluster with overall bad system performance.

We can now interact with the data by applying one or more filters on any desired axis. Figure 6 shows a screenshot of the parallel coordinate plot after filters were applied on all experiments that end in cluster 1 and 2. This helps for a better visibility of the distribution of parameter values in those clusters. Furthermore, the input factor interarrival time was added.

Now we can analyze how the values of this factor are distributed among the highlighted clusters. For the good performance cluster (blue), only small values can be found. Therefore, we can conclude that short interarrival times presumably lead to a good system

performance. On the other hand, we cannot conclude that large interarrival times in turn lead to a bad system performance because the factor values of interarrival time are equally distributed among the bad performance cluster (cyan).

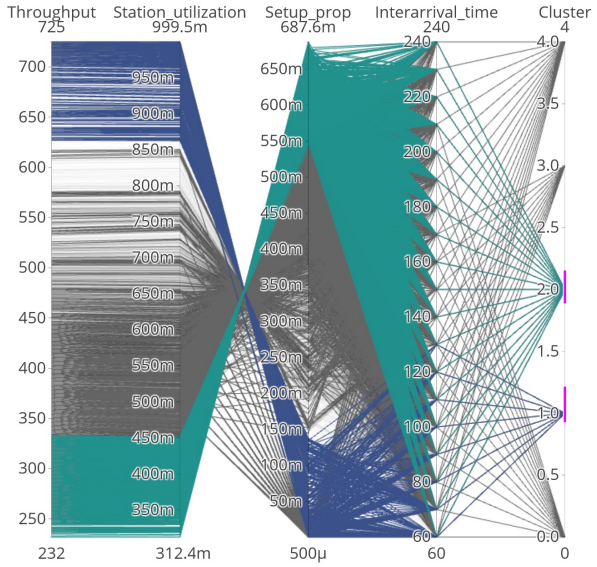

Figure 6. Filter for cluster 1 and 2 is active and the factor 'interarrival time' was added. Corresponding experiments in these clusters can be traced more accurately.

We can then add more factors to the parallel coordinates plot in order to investigate their impact on the cluster allocation. Figure 7 shows a screenshot after the factor sorter capacity has been added additionally.

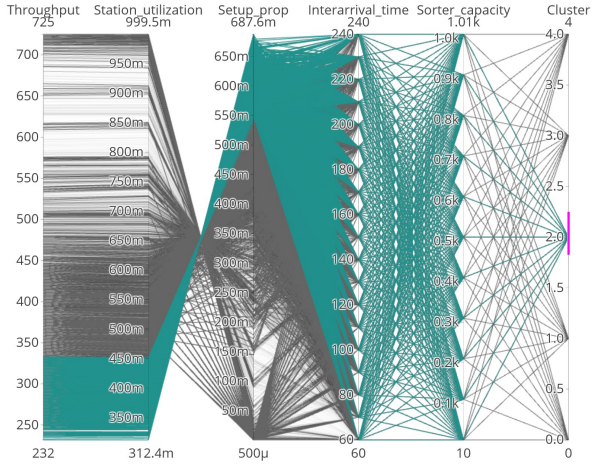

Figure 7. Additionally, the factor 'sorter capacity' was added to the plot. We can see that it distributes equally among cluster 2.

We can see that sorter capacity does also not affect the cluster allocation at all because the factor values are equally distributed. This is not only true for cluster 2 but for all of the five clusters. This effect becomes apparent when interactively sliding the filter on the cluster axis. This will change the coloration but not the distribution on the sorter capacity axis. In other words, there is no difference in the distribution of sorter capacity between all generated clusters.

In order to investigate the factor sorting strategy, we need to switch to a visualization that is suitable for categorical parameters. For this purpose, we generated two bar charts showing the distribution of the sorting strategies filtered for the good performance cluster and bad performance cluster. This visualization is shown in Figure 8. We can clearly see that the setup optimal sorting strategy is very dominant in cluster 1, whereas it is completely absent in cluster 2. Therefore, we can conclude that the sorting strategy does also have a strong impact on system performance.

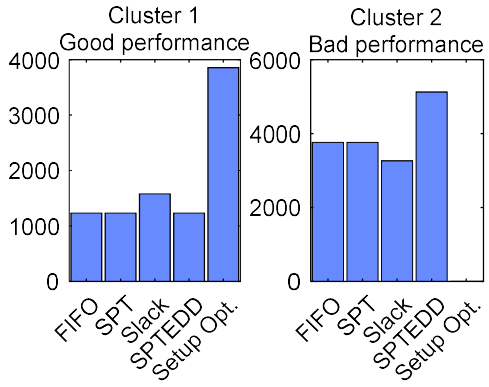

Figure 8. Bar charts for comparison of cluster 1 vs. cluster 2 concerning factor 'sorting strategy'.

Presumably, there might be even an interaction between interarrival time and sorting strategy. To investigate this, we used scatterplots with corresponding regression lines on subsets that were filtered by sorting strategies. This is shown in Figure 9 in terms of a comparison between FIFO and the setup optimal sorting strategy.

Firstly, we can see the direct negative effect that an increasing interarrival time has on the throughput in both plots. When going from FIFO to setup optimal sorting, the magnitude of the effect (slope of the regression line) is much higher. We can therefore conclude that setup optimal sorting leverages the effect on throughput even more. In addition, we can see that there are much more experiments (black dots) in the

area of throughput > 600 in the right plot compared to the left.

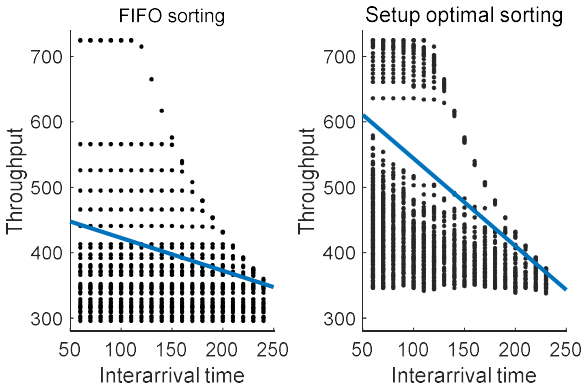

Figure 9. Interaction plots for interarrival time versus throughput, filtered by different sorting strategies (FIFO and setup optimal sorting). The blue line represents a linear regression between interarrival time and throughput.

In summary, we can therefore conclude that short interarrival time is the basic prerequisite for good system performance. When the load on the system is high, this effect reverts itself because the station is bottlenecking when lots of setup processes need to be performed. The use of setup optimal sorting reduces the bottleneck in order to increase the throughput from this point. In further analyses, we could now investigate the exact location of the break-even point between high throughput and bottlenecking. Furthermore, the impact of product mixture regarding individual product proportions could be investigated.

Despite hidden relations are rather limited and findings are obvious to some extent given a single server model, we were able to successfully create knowledge about the system using an interactive, visually aided analysis.

5. Conclusion and Future Work

In this paper, we showed how the process of knowledge discovery in simulation can be applied using interactive visual analysis. For this purpose, we compiled a list of suitable visualization methods for given analysis operations. We then showed how visualization tasks can be composed of typical data farming analysis objectives alongside suitable visualization and interaction methods. We demonstrated the application of this process by means of a simple extended single server model.

Future research is needed for creating visualization methods and toolsets that are especially suitable for visual analytics in the context of manufacturing simulation data, or even discrete event simulation data in general. For the demonstrations shown in this paper, we used a prototype that was composed of Apache Spark, MatLab and various JavaScript Frameworks. For visual analytics to become commonplace, an integration of the demonstrated methods with commonly applied simulation packages is crucial. By talking to practitioners, we found out that this approach to simulation data analysis based on visualization and interaction with a human in the loop is more motivating and appealing to non-simulation experts than traditional approaches. This yields research possibilities regarding usability aspects to improve cognitive perception of visualizations. Additionally using stream-based data mining algorithms could be used for real time analysis of simulation data, even when experimentation has not been finished completely. First investigations of this approach yield very promising results [9]. In order to support an online analysis of simulation experiments visually, visualization methods that are able to build up iteratively in accordance with the flow of data are required.

6. References

[1] Behrisch, M., D. Streeb, F. Stoffel, D. Seebacher, B. Matejek, S.H. Weber, S. Mittelstaedt, H. Pfister, and D. Keim, "Commercial Visual Analytics Systems - Advances in the Big Data Analytics Field", IEEE Transactions on Visualization and Computer Graphics, 2018.

[2] Brehmer, M. and T. Munzner, "A Multi-Level Typology of Abstract Visualization Tasks", IEEE Transactions on Visualization and Computer Graphics, 19(12), 2013, pp. 2376–2385.

[3] Choo, J., S. Bohn, and H. Park, "Two-stage Framework for Visualization of Clustered High Dimensional Data", in 2009 IEEE Symposium on Visual Analytics Science and Technology (VAST), 2009 IEEE Symposium on Visual Analytics Science and Technology, Atlantic City, NJ, USA, 12.10.2009 - 13.10.2009. 2009. IEEE.

[4] Choo, J. and H. Park, "Customizing Computational Methods for Visual Analytics with Big Data", IEEE Computer Graphics and Applications, 33(4), 2013, pp. 22–28.

[5] Cioppa, T.M. and T.W. Lucas, "Efficient Nearly Orthogonal and Space-Filling Latin Hypercubes", Technometrics, 49(1), 2007, pp. 45–55.

[6] Elmegreen, B.G., S.M. Sanchez, and A.S. Szalay, "The Future of Computerized Decision Making", in Proceedings of

the 2014 Winter Simulation Conference, A. Tolk, S.D. Diallo, I.O. Ryzhov, L. Yilmaz, S. Buckley, and J.A. Miller, Editors, 2014 Winter Simulation Conference, Savannah GA, 7.12.-10.12. 2014. IEEE Inc: Piscataway, N.J.

[7] Fayyad, U.M., G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases", AI Magazine, 17, 1996, pp. 37–54.

[8] Feldkamp, N., S. Bergmann, and S. Strassburger, "Knowledge Discovery in Manufacturing Simulations", in Proceedings of the 3rd ACM SIGSIM Conference on Principles of Advanced Discrete Simulation, S.J.E. Taylor, N. Mustafee, and Y.-J. Son, Editors. 2015. ACM: New York, NY, USA.

[9] Feldkamp, N., S. Bergmann, and S. Strassburger, "Online Analysis of Simulation Data with Stream-based Data Mining", in Proceedings of the 2017 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation - SIGSIM-PADS '17, W. Cai, T.Y. Meng, P. Wilsey, and K. Jin, Editors, the 2017 ACM SIGSIM Conference, Singapore, Republic of Singapore, 24-26.05. 2017. ACM Press: New York, New York, USA.

[10] Feldkamp, N., S. Bergmann, S. Strassburger, and T. Schulze, "Knowledge Discovery in Simulation Data: A Case Study of a Gold Mining Facility", in Proceedings of the 2016 Winter Simulation Conference, T.M.K. Roeder, P.I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S.E. Chick, Editors, Winter Simulation Conference (WSC), Washington, DC, USA, 11-14.12. 2016. IEEE Inc: Piscataway, N.J.

[11] Feldkamp, N., S. Bergmann, S. Strassburger, T. Schulze, P. Akondi, and M. Lemessi, "Knowledge Discovery in Simulation Data – a Case Study for a Backhoe Assembly Line", in Proceedings of the 2017 Winter Simulation Conference, V. Chan, A. D'Ambrogio, G. Zacharewicz, and N. Mustafee, Editors, Las Vegas, 3-6.12. 2017. IEEE Inc.

[12] Han, J. and M. Kamber, Data mining: Concepts and techniques, 2nd edn., Elsevier; Morgan Kaufmann, Amsterdam, Boston, San Francisco, CA, 2006.

[13] Hartigan, J.A., "Printer Graphics for Clustering", Journal of Statistical Computing and Simulation, 4, 1975, pp. 187–213.

[14] Horne, G., B. Åkesson, T. Meyer, and S. Anderson, Data farming in support of NATO: Final Report of Task Group MSG-088, North Atlantic Treaty Organisation, Neuilly-sur-Seine Cedex, 2014.

[15] Horne, G.E. and T.E. Meyer, "Data Farming: Discovering Surprise", in Proceedings of the 2005 Winter Simulation Conference, M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Joines, Editors, Orlando, FL. USA, 4.12. 2005. IEEE Inc: Piscataway, N.J.

[16] Inselberg, A., "The plane with parallel coordinates", The Visual Computer, 2(1), 1985, pp. 69–91.

[17] Keim, D., G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, "Visual Analytics: Definition, Process, and Challenges", in Information Visualization: Human-Centered Issues and Perspectives, A. Kerren, J.T. Stasko, J.-D. Fekete, and C. North, Editors. 2008. Springer Berlin Heidelberg: Berlin, Heidelberg.

[18] Keim, D.A., "Information Visualization and Visual Data Mining", IEEE Transactions on Visualization and Computer Graphics, 8(1), 2002, pp. 1–8.

[19] Keim, D.A., F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler, "Visual Analytics: Scope and Challenges", in Visual Data Mining: Theory, Techniques and Tools for Visual Analytics, S. Simoff, M.H. Boehlen, and A. Mazeika, Editors. 2008. Springer: Berlin, Heidelberg.

[20] Kennedy, P.J., S.J. Simoff, D.R. Catchpoole, D.B. Skillicorn, F. Ubaudi, and A. Al-Oqaily, "Integrative Visual Data Mining of Biomedical Data: Investigating Cases in Chronic Fatigue Syndrome and Acute Lymphoblastic Leukaemia", in Visual Data Mining: Theory, Techniques and Tools for Visual Analytics, S. Simoff, M.H. Boehlen, and A. Mazeika, Editors. 2008. Springer: Berlin, Heidelberg.

[21] Kleijnen, J.P.C., S.M. Sanchez, T.W. Lucas, and T.M. Cioppa, "State-of-the-Art Review: A User's Guide to the Brave New World of Designing Simulation Experiments", INFORMS Journal on Computing, 17(3), 2005, pp. 263–289.

[22] Law, A.M., Simulation Modeling and Analysis, 5th edn., McGraw Hill Book Co, New York, N.Y., 2014.

[23] Luboschik, M., C. Tominski, A. Bittig, A. Uhrmacher, and H. Schumann, "Towards Interactive Visual Analysis of Microscopic-Level Simulation Data", in Proceedings of SIGRAD 2012: Interactive Visual Analysis of Data, A. Kerren and S. Seipel, Editors, 29.-30.11.2012. 2012. Linköping University Electronic Press: Linköpings.

[24] Matkovic, K., D. Gracanin, M. Jelović, and H. Hauser, "Interactive Visual Analysis of Large Simulation Ensembles", in Proceedings of the 2015 Winter Simulation Conference, L. Yilmaz, W.K.V. Chan, I. Moon, T.M.K. Roeder, C. Macal, and M.D. Rossetti, Editors, 2015 Winter Simulation Conference, Huntington Beach, 07-09.12. 2015. IEEE Inc: Piscataway, N.J.

[25] Meyer, T.E. and S.K. Johnson, "Visualization for Data Farming: A Survey of Methods", in Maneuver Warfare Science 2001, G.E. Horne and M. Leonardi, Editors. 2001. Marine Corps Combat Development Command: Quantico, Virginia, USA.

[26] Puolamäki, K., A. Bertone, R. Therón, O. Huisman, J. Johansson, S. Miksch, P. Papapetrou, and S. Rinzivillo, "Data Mining", in Mastering the information age: Solving problems with visual analytics, D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, Editors. 2010. Eurographics Association: Goslar.

[27] Reiterer, H. and H.-C. Jetter, "Informationsvisualisierung", in Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis, R. Kuhlen, W. Semar, and D. Strauch, Editors. 2013. Walter de Gruyter: Berlin.

[28] Risch, J., A. Kao, S.R. Poteet, and J. Wu, "Text Visualization for Visual Text Analytics", in Visual Data Mining: Theory, Techniques and Tools for Visual Analytics, S. Simoff, M.H. Boehlen, and A. Mazeika, Editors. 2008. Springer: Berlin, Heidelberg.

[29] Sanchez, S. and P.J. Sanchez, "Better Big Data via Data Farming Experiments", in Advances in Modeling and Simulation: Seminal Research from 50 Years of Winter Simulation Conferences, A. Tolk, J. Fowler, G. Shao, and E. Yücesan, Editors. 2017. Springer International Publishing.

[30] Sanchez, S.M., "Work Smarter, Not Harder: Guidelines for Designing Simulation Experiments", in Proceedings of the 2007 Winter Simulation Conference: December 9 - 12, 2007, Washington, DC, U.S.A, S.G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J.D. Tew, and R.R. Barton, Editors. 2007. IEEE: Piscataway, N.J.

[31] Sanchez, S.M., "Simulation Experiments: Better Data, Not Just Big Data", in Proceedings of the 2014 Winter Simulation Conference, A. Tolk, S.D. Diallo, I.O. Ryzhov, L. Yilmaz, S. Buckley, and J.A. Miller, Editors, 2014 Winter Simulation Conference, Savannah GA, 7.12.-10.12. 2014. IEEE Inc: Piscataway, N.J.

[32] Schubert, J., R. Johansson, and P. Hörling, "Skewed Distribution Analysis in Simulation-Based Operation Planning", in Ninth Operations Research and Analysis Conference, N. Carson and A. Williams, Editors, Ottobrunn, Germany, 22-23.10. 2015.

[33] Soban, D., D. Thornhill, S. Salunkhe, and A. Long, "Visual Analytics as an Enabler for Manufacturing Process Decision-making", Procedia CIRP, 56, 2016, pp. 209–214.

[34] Steed, C.A., K.J. Evans, J.F. Harney, B.C. Jewell, G. Shipman, B.E. Smith, P.E. Thornton, and D.N. Williams, "Web-based Visual Analytics for Extreme Scale Climate Science", in 2014 IEEE International Conference on Big Data, 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 27.10.2014 - 30.10.2014. 2014. IEEE.

[35] Strassburger, S., S. Bergmann, N. Feldkamp, K. Sokoll, and M. Clausing, "Data Farming Research Project with Audi and VW", in 2018 Plant Simulation Worldwide User Conference, Siemens AG, Editor, 2018 Plant Simulation Worldwide User Conference, Stuttgart, 16.-18.10.2018. 2018.

[36] Thomas, J.J. and K.A. Cook, Illuminating the Path: Research and Development Agenda for Visual Analytics, 1st edn., IEEE Computer Society, Los Alamitos, California, 2005.

[37] Vieira, H., S.M. Sanchez, K.H. Kienitz, and M.C.N. Belderrain, "Improved efficient, nearly orthogonal, nearly balanced mixed designs", in Proceedings of the 2011 Winter Simulation Conference, S. Jain, R. Creasey, J. Himmelspach, K.P. White, and M.C. Fu, Editors, Phoenix, AZ, USA, December 11 - 14, 2011. 2011. IEEE Inc: Piscataway, NJ.

[38] Wegman, E.J., "Hyperdimensional data analysis using parallel coordinates", Journal of the American Statistical Association, 85(441), 1990, pp. 664–675.

[39] Wei, J., Z. Shen, N. Sundaresan, and K.-L. Ma, "Visual Cluster Exploration of Web Clickstream Data", in 2012 IEEE Conference on Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on Visual Analytics Science and Technology (VAST), Seattle, WA, USA, 14.10.2012 - 19.10.2012. 2012. IEEE.

[40] Wenzel, S., J. Bernhard, and U. Jessen, "Visualization for modeling and simulation: a taxonomy of visualization techniques for simulation in production and logistics", in Proceedings of the 2003 Winter Simulation Conference, S. Chick, P.J. Sanchez, D. Ferrin, and D.J. Morrice, Editors, New Orleans, LA, USA, 7-10.12. 2003. IEEE inc.: Piscataway, N.J.